More on Weighting Factors for Complicated Models

P.G. Sullivan[1], Z. Liu[2], J.H. Jakobsen[3] and W.F. Fikse[3]
[1]CDN, Canada, [2]VIT, Germany, [3]Interbull Center, Sweden

Introduction

Current methods to derive weighting factors (effective daughter contributions, or EDC) for MACE can be applied to a wide range of national evaluation models, but are in many cases sub-optimal. Improvements suggested by Sullivan *et al.* (2004) corrected major errors in EDC for sire models and repeatability models, and extended the methods to models with direct and maternal effects, threshold models and data sets that include performance records for both sires and their progeny. However, problems with the EDC for multiple-trait models, as suggested by Liu *et al.* (2001, 2002), were not addressed by Sullivan *et al.* (2004).

Both the current (EDC2000) and proposed methods (EDC2004) of Sullivan *et al.* (2004) accumulate multiple-trait progeny information into sire EDCs based on single-trait absorption equations, overestimating contributions from correlated traits of multiple progeny. For example, reliability from only a correlated trait approaches 100% in the limit with these methods, instead of the correct value which is the square of the genetic correlation.

Liu *et al.* (2001) outlined procedures to more correctly accumulate multiple-trait progeny information (MT-EDC), and Jakobsen *et al.* (2003) extended MT-EDC to a model with direct and maternal genetic effects. Testing of MT-EDC has so far been limited to very small data sets, however, and a general-purpose program to compute MT-EDC for a wide array of national models is not yet available. Therefore, the purpose of this study was to further develop and test suitable methods and software to derive EDC under the general class of linear models, which may include multiple traits and/or genetic effects, and genetic random regressions.

Methods

The basic principle behind MT-EDC was to accumulate progeny information, for multiple-trait or multiple-genetic-effect models, into sire EDC matrices instead of EDC scalar quantities for each trait. The derivation was based on the absorption of mixed-model equations of progeny and mates into the sire, and involved repeated conversions between reliability and information matrices of individuals. The conversions to reliability matrices were useful for comparisons with the EDC2000 methods, but are not required to properly accumulate progeny information into multivariate EDC matrices of sires. For the present study, a new multi-trait algorithm was developed that did not include repeatedly generating reliability matrices.

As is common for EDC and approximate reliability algorithms, estimation of all but one of the fixed factors is ignored. The basic approach is to absorb equations for the main fixed effect (e.g. contemporary group, CG) into records of animals, then perform a series of carefully ordered absorptions of equations among animals, as follows:

Step 1) Absorb CG equations into animals, accounting for paternal sib relationships within CG. This step is done separately for each trait (e.g. Interbull, 2000), and the paternal sib relationships are specific to the genetic effect in the model (common sires for direct effects of animals and common maternal grandsires for maternal effects of dams).

Step 2) Effective multi-trait information, with CG absorbed, is inferred from the effective records for each trait, as described in Appendix 1, and stored for each animal in a vector NE.

Step 3) Absorb NE of recipients (maternal dams) into NE of ET animals.

Step 4) Make a copy of NE into NE2 for each animal.

Step 5) Chronologically (progeny before parents) absorb NE of progeny into genetic sire and dam; a) into NE while assuming mates of each parent are unknown, and b) into NE2 while recognizing whether or not the mates are known. The unknown-mates assumption for NE avoids major difficulties associated with absorption of mate information when generations overlap.

Step 6) Absorb NE2 of dams into NE2 of sires, accounting for the specific combinations of full-sib and half-sib progeny of each mating pair. This step is done one sire at a time, using a list of progeny of the sire, ordered within mates, to properly accumulate absorptions of progeny NE into the sire X dam equations, as needed to absorb dam into sire.

Step 7) From NE2, derive EDC for all traits of interest for MACE, which may be any set of linear functions of the genetic effects in the national model (Appendix 2), and could also include EDC matrices for multiple-trait MACE applications.

Application and Testing

The above algorithm was added to the general EDC program described by Sullivan *et al.* (2004), which was designed to accommodate models with any combination of single or multiple traits, with or without maternal effects for all or a subset of the traits included. Models may also include permanent environmental variances for animals and/or dams.

Data were created following the general simulation described by Sullivan *et al.* (2004), with a few improvements to allow better comparisons among EDC alternatives. Sires were randomly assigned to have completely missing information for some traits, the average progeny count per sire was increased from 10 to 75 and the variation of progeny counts was increased by adding a fourth category for the frequency of sire usage (mating probabilities of 4:3:2:1 among the four sire groups). Additional features of the simulation included variation among CG sizes and sires per CG, combinations of natural and ET progeny records, and variation in record weightings (e.g. for models with heterogeneous residual variances, combinations of complete and projected lactation records, etc.).

From the simulated data, several input data sets were created to compute EDCs under a variety of models, a subset of which are presented in this paper. Approximate reliabilities, derived from the inverse of mixed-model equations based on EDC, were compared against true reliabilities, derived from the inverse of mixed-model equations based on performance records. The approximate reliabilities should be highly correlated with, and have similar means and standard deviations compared to, the true reliabilities. Ideally there should be consistent agreement between true and approximate reliabilities for all types of models, which is critical to minimize potentially spurious effects of national model types on the rankings of countries in MACE results.

Three alternatives to the current EDC methods, labelled ITB, were compared: The methods of Sullivan *et al.* (2004), labelled ST, which are based on single-trait absorption of progeny information, the methods of Jakobsen *et al.* (2002), labelled MT, which are based on multiple-trait absorption of progeny information, and the general mixed model approach, labelled MM, which combines desirable features from each of ST and MT.

Results are shown in Table 1 for the latter 3 methods, ITB excluded as it does not accommodate models with maternal effects (e.g. calving ease). Reliabilities based on MM-EDC were closest to the true reliabilities for both sires and mates of the females in the simulation, who had both own and progeny performance records. Correlations with true reliabilities were highest, and errors in the first two moments (u and s) were generally lowest with MM-EDC, for both direct and maternal genetic effects. The variation among animals and effects, in the agreement between approximate and true reliabilities, was also much lower for MM than either ST or MT.

For sires of females, having a combination of progeny and grand-progeny performance records in the data, reliabilities were underestimated for direct effects by ST and MT, which ignored grand-progeny records completely, and overestimated by MM, which accounted for grand-progeny records but ignored information about grand-progeny mates. Because this was a single-trait model, there were few if any practical advantages for MT over ST. The advantages of MM can be attributed to a combination of improved adjustments for mates, sibs and grand-progeny.

Table 1. Squared correlation (R^2) between approximate and true reliabilities, and relative difference ((approximate-true)/true) in means (u) and standard deviations (s) of reliabilities for three different weighting factors (ST, MT and MM defined in the text) applied to a single-trait animal model with direct and maternal genetic effects and with or without contemporary group (Xb) effects (y= μ +Xb+Za or y= μ +Za, direct h^2=.05, maternal h^2=.05, direct-maternal genetic correlation = -.1 and maternal repeatability = .20).

		100 * R^2			u			s			True
Effect	**Xb**	**ST**	**MT**	**MM**	**ST**	**MT**	**MM**	**ST**	**MT**	**MM**	**Rel (%)**
***** SIRES OF FEMALES (n=12) *****											
direct	**no**	100	100	100	-11	-14	8	-6	-13	16	21 ± 9
maternal	**no**	100	100	100	4	12	4	5	9	4	29 ± 10
direct	**yes**	100	99	100	-9	-12	4	-7	-13	10	20 ± 8
maternal	**yes**	100	100	100	8	20	10	11	18	11	25 ± 9
***** MATES OF FEMALES (n=12) *****											
direct	**no**	100	100	100	3	3	3	11	10	4	34 ± 11
maternal	**no**	96	100	100	72	9	2	176	8	5	0.3 ± 0.1
direct	**yes**	100	96	100	6	-1	3	11	9	3	26 ± 8
maternal	**yes**	98	96	100	51	7	2	110	10	2	0.3 ± 0.1

Table 2. Squared correlation (R^2) between approximate and true reliabilities, and relative difference ((approximate-true)/true) in means (u) and standard deviations (s) of reliabilities for three different weighting factors (ITB, ST and MM defined in the text) applied to a 2-trait animal model with no contemporary group (Xb) effects (y= μ +Za, h^2=.25 for both traits).

			100 * R^2			u			s			True
Trait	**r_G**	**r_E**	**ITB**	**ST**	**MM**	**ITB**	**ST**	**MM**	**ITB**	**ST**	**MM**	**Rel (%)**
***** MATES OF FEMALES (n=12) *****												
1	**.00**	**.00**	100	100	100	3	3	2	7	7	4	70 ± 11
1	**.70**	**.00**	99	99	100	4	4	2	9	9	4	71 ± 10
1	**.70**	**.40**	100	100	100	3	3	1	9	8	5	71 ± 10
1	**.90**	**.00**	100	100	100	4	4	2	10	10	5	72 ± 10
1	**.90**	**.70**	97	96	96	1	0	0	10	9	6	73 ± 10
2	**.00**	**.00**	100	100	100	11	11	10	12	12	11	30 ± 32
2	**.70**	**.00**	78	78	100	29	29	4	-13	-13	10	50 ± 19
2	**.70**	**.40**	73	72	100	28	27	4	-16	-17	9	50 ± 18
2	**.90**	**.00**	92	92	100	13	13	3	5	5	7	64 ± 12
2	**.90**	**.70**	83	79	98	10	9	0	1	-1	3	64 ± 12
ave	**.00**	**.00**	75	75	100	32	32	6	-22	-22	10	43 ± 23
ave	**.70**	**.00**	88	88	100	12	12	4	-3	-3	8	64 ± 13
ave	**.70**	**.40**	86	84	100	12	11	3	-1	-2	7	63 ± 13
ave	**.90**	**.00**	98	98	100	6	6	3	6	6	6	70 ± 11
ave	**.90**	**.70**	92	90	97	4	3	0	6	5	4	70 ± 11

Table 3. Squared correlation (R^2) between approximate and true reliabilities, and relative difference ((approximate-true)/true) in means (u) and standard deviations (s) of reliabilities for three different weighting factors (ITB, ST and MM defined in the text) applied to a 2-trait animal model with contemporary group (Xb) effects (y= μ +Xb+Za, h^2=.25 for both traits). Average contemporary group sizes were 7.0 and 2.3 for traits 1 and 2 respectively.

			100 * R^2			u			s			True
Trait	**r_G**	**r_E**	**ITB**	**ST**	**MM**	**ITB**	**ST**	**MM**	**ITB**	**ST**	**MM**	**Rel (%)**
***** MATES OF FEMALES (n=12) *****												
1	**.00**	**.00**	100	100	100	8	8	8	14	14	11	60 ± 9
1	**.70**	**.00**	100	100	100	9	9	8	16	16	11	61 ± 9
1	**.70**	**.40**	100	100	100	9	8	8	15	14	11	60 ± 9
1	**.90**	**.00**	100	100	100	9	9	8	16	16	12	62 ± 9
1	**.90**	**.70**	99	100	100	9	8	7	16	14	12	61 ± 9
2	**.00**	**.00**	100	100	100	16	16	15	17	17	16	20 ± 22
2	**.70**	**.00**	83	83	100	37	37	9	4	4	15	40 ± 13
2	**.70**	**.40**	81	76	100	36	34	9	3	-1	14	40 ± 13
2	**.90**	**.00**	95	95	100	19	19	8	18	18	13	54 ± 10
2	**.90**	**.70**	95	91	100	19	17	8	21	17	12	53 ± 9
ave	**.00**	**.00**	79	79	100	38	38	11	-11	-11	16	34 ± 16
ave	**.70**	**.00**	91	91	100	18	18	8	9	9	14	54 ± 11
ave	**.70**	**.40**	92	88	100	19	17	9	13	10	15	53 ± 10
ave	**.90**	**.00**	99	99	100	12	12	8	14	14	12	60 ± 10
ave	**.90**	**.70**	99	98	100	12	11	7	18	15	12	58 ± 9

Results for a number of 2-trait models are in Tables 2 and 3. Method MT was not applied to these models due to limited time and a lack of available software. However, given the observations for a single-trait model, and the theoretical differences among methods, MT was expected to be intermediate between ST and MM for multiple-trait models. MT uses multi-trait absorptions of progeny but does not consider grand-progeny or maternal sib information.

Correlations between approximate and true reliabilities, for trait 2 with a non-zero genetic correlation, were much higher for MM relative to ITB and ST, and the average biases in mean reliability were much lower. Biases seemed higher for trait 2 versus 1 when correlation parameters were zero, but mainly because bias was expressed as a percentage of much lower reliabilities for trait 2. The standard deviation of approximate reliabilities was always biased upwards for MM. The variation among models, in biases of means and standard deviations, was lowest for MM. Performance of all three methods, for the average (a linear function) of 2 traits was generally intermediate between the respective performances for the individual traits.

When contemporary group was added to the model (Table 3), biases increased for all methods while the ranking of methods did not change. MM gave higher correlations with true values and lower biases in approximated reliabilities. The biases from imperfect absorption of CG (Table 3) were larger than biases attributed to imperfect absorption of genetic effects with MM (Table 2), suggesting that greater opportunities for improving the methods further are likely in the absorption of CG effects.

For trait 1 in 2-trait models, differences between observations for sires (not shown), relative to mates of females were similar to corresponding differences for the single-trait model (Table 1). For a genetically correlated trait 2, ITB and ST performed better for sires than for mates of females, because the upward biases for trait 2 were offset by downward biases from ignoring grand-progeny. Otherwise, MM was consistently the best method for sires of females for all models.

Conclusions

Methods and software to compute EDCs for MACE evaluations were extended to incorporate multivariate absorption techniques under the general class of linear models. The software should be tested with field data from countries participating in MACE evaluations. Use of a common EDC program could efficiently harmonize EDC calculations among countries and improve MACE results.

References

Interbull, 2000. New weighting factors for the international genetic evaluation; revised July, 2000. Mimeo. (OR see current document, www-interbull.hgen.slu.se)

Jakobsen, J.H., Fikse, W.F., Liu, Z. & Sullivan, P.G. 2003. Outline of a procedure to calculate weighting factors for models with maternal effects. *Interbull Bulletin 31,* 65-69.

Liu, Z, Reinhardt, F & Reents, R. 2001. The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values. *Interbull Bulletin 27,* 41-47.

Liu, Z., Reinhardt, F. & Reents, R. 2002. The multiple trait effective daughter contribution method applied to approximate reliabilities of EBVs of a random regression test day model for genetic evaluation in dairy cattle. *Proc 7th WCGALP* comm. #20-15.

Sullivan, P.G, Kistemaker, G.J. Jakobsen, J & Fikse, F. 2004. MACE weighting factors for direct and maternal EBVs from multiple-trait models. *Interbull Bulletin 32,* 53-58.

APPENDIX 1
Inferring combinations of records

Let **n** be a vector of squared observation counts (Z'Z) for a given animal, with n_i representing the number of observations for trait/effect *i*. The number of squared observations for a combination that includes all traits with n_i >= min(**n**) is min(**n**). Subtracting min(**n**) from all n_i and repeating gives the number for each subsequent trait combination until all n_i are 0. Matrix $Z'R^{-1}Z$ can then be computed for the animal by summing products of record combination frequencies and corresponding R^{-1} matrices. Matrix R should ideally be with respect to the given individual (i.e. residual covariances between traits are generally higher within animal than they are within sire).

APPENDIX 2
Deriving EDC of linear functions

Given a matrix of information, for a given individual α, on a set of traits or genetic effects ($\mathbf{Z_\alpha'MZ_\alpha}$) and a corresponding genetic covariance matrix (**G**), the *EDC* for a linear function of the given effects (**L'g**), with sire variance ratio *k,* is derived as follows:

$$\text{Let } \mathbf{C} = [\mathbf{Z_\alpha'MZ_\alpha} + \mathbf{G}^{-1}]^{-1}$$
$$R = 1 - \mathbf{L'CL/L'GL}$$
$$EDC = kR/(1-R)$$